Automatic Emotion Recognition Using Hybrid Feature Extraction And DEEP Learning With EEG Signals

Dr. A. Ezil Sam Leni[1], S. Prasanth[2], R. Sanjay[3]
[1,] HOD, Dept of Computer Science
[2, 3]Dept of Computer Science
[1, 2, 3] Jeppiaar SRR Engineering College.

Abstract *- Automatic emotion recognition is one of the most challenging tasks. It is an important task for computer to understand the human status in Brain Computer Interface (BCI) systems. Recently, recognition of human emotions via Physiological signals or Electroencephalogram (EEG) signals has become an active research topic and researchers strive for the advancement of techniques, algorithms and methodologies used in it. The typical workflow of a classification problem involves Data collection, Pre-Processing, Feature extraction, Feature selection and Classification. In this work, we use pre-processed SEED dataset as EEG data source, Hybrid Feature Extraction technique for feature extraction, Fisher Scoring algorithm for feature selection and Deep Learning Neural Network for classification.*

Keywords- EEG, EEG Emotion, Emotion Detection, Fisher Score feature selection, Hybrid Feature Extraction, Neural Network Classification.

I. INTRODUCTION

When a person does or experience any physical or mental activities (including experienced emotions), the neurons in his / her brain fires up minute voltages forming specific patterns for each class of emotion. This voltage fluctuation can be recorded with the help of EEG headsets in the form of Electroencephalogram (EEG) signal.

Basically, Emotion of a person can be classified into two types: Expressed emotion (or Shown emotion), and Experienced Emotion (or Felt emotion). Expressed emotion of a person can be identified by his / her physical activities such as facial expressions, tone of speech, attitude and behavior. Internally felt emotion of a person is impossible to find using physical characteristics of that person, unless expressed. Hence, EEG signals recorded from that person's brain using EEG headsets are used to identify Internally felt emotions. The main focus of this work is to classify the EEG signals recorded from a person using a EEG Headset into three categories: Positive Emotion, Neutral Emotion, and Negative Emotion.

II. EEG DATA COLLECTION

SEED dataset[1] is used as data source for this work. This dataset includes EEG signals recorded from 15 subjects, while they were watching audio-visual media, which stimulates emotions. In order to investigate neural signatures and stable patterns across sessions and individuals, each subject was required to perform the experiments for three sessions. Hence there are totally 45 samples in this dataset.

The EEG Headset used to record these signals is a 62-channel active AgCl electrode cap with electrodes positioned according to International 10-20 standard.

Preprocessing includes 200 Hz Down-Sampling, 0-75 Hz Band Pass Filtering and segment of EEG signal while the video media was running was separated.

III. FEATURE EXTRACTION

Hybrid Feature Extraction [2]involves features derived using Discrete Wavelet Transform (DWT), Wavelet Packet Transform (WPT), Auto Regression (AR), and Raw EEG signal, and fusing them altogether as a complete the list of features.

The baseline features include Max of Absolute values, Mean of Absolute values, Shannon Entropy, Power, Relative Power, Root Mean Square, Standard Deviation and Variance. These features are extracted after every individual signal processing method in Hybrid Feature Extraction.

1. Discrete Wavelet Transform (DWT)

In wavelet analysis, the Discrete Wavelet Transform (DWT) [4] decomposes a signal into a set of mutually orthogonal wavelet basis functions. These functions differ from sinusoidal basis functions in that they are spatially localized – that is, nonzero over only part of the total signal length. Furthermore, wavelet functions are dilated, translated and scaled versions of a common function φ, known as the mother wavelet. As is the case in Fourier analysis, the DWT is invertible, so that the original signal can be completely recovered from its DWT representation.

In this work, 'Daubechies 8' or 'db8' is used as mother wavelet and the signal is decomposed into 5 levels.

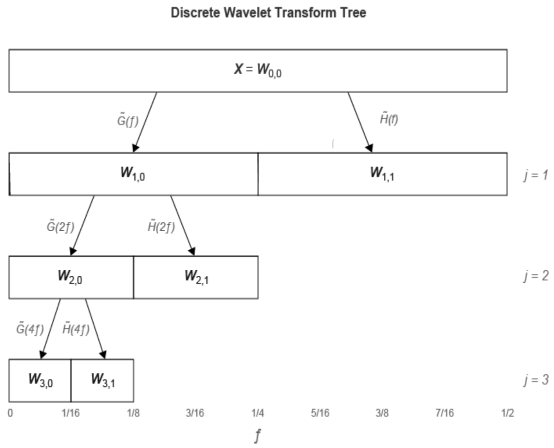

Fig. 1DWT decomposition into 3 levels

2. Wavelet Packet Transform (WPT)

The WPT of a signal generates packets of coefficientscalculated by scaling and shifting a chosen mother wavelet, which is a prototype function. Accordingly, at the 1st level of the WPT, the original signal S is split into two frequency band packets called approximation, A1, and detail, D1. At the 2nd level, each approximation and detail packet are again split into further approximations, AA2 and AD2, and details, DA2 and DD2, and the process is repeated in the next levels generating other decomposition packets.

'Daubechies 8' or 'db8' is selected as mother wavelet and the signal is decomposed into 3 levels.

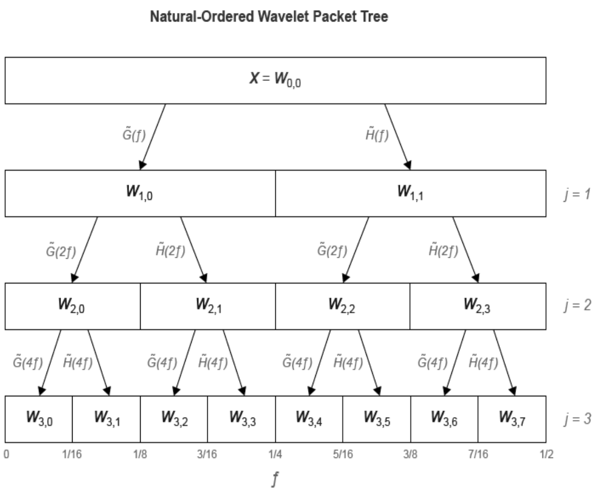

Fig. 2WPT decomposition into 3 levels

3. Auto Regression (AR)

The AR model has high ability in representing and modelling the characteristics and information inside a signal. AR model is frequently used in different approaches toward processing EEG signals such as: BCI designs, classification of schizophrenic patients, estimation of hypnosis levels, determination of sleep stages, analysis of anaesthesia, classification of epilepsy diagnosis, and for emotion detection. In AR model, each sample is obtained from the summation of previous weighted samples according to (1). The model order is determined by the number of weights, which are called AR coefficients.

$$x(t) = -\sum_{i=1}^{p} a_i x(t-i) \tag{1}$$

Where, P is the model order and AR coefficients are denoted as a_i (i = 1,…,p). In this paper, AR coefficients are obtained by applying Burg method. In Burg method, AR reflection coefficients are estimated by minimizing the sum of forward and backward forecasted errors.

k_p is The p^{th} reflection coefficient which is a criterion of the correlation between $x(t)$ and $x(t-p)$. By applying the Levinson–Durbin recursion algorithm, these reflection coefficients k_i can be converted, into AR parameters according to (2)

$$a_{p,i} = \begin{cases} a_{p-1,i} + k_p a^*_{p-1,p-i} & i = 1,\ldots,p-1 \\ k_p & i = p \end{cases} \tag{2}$$

In the p^{th} level of Burg method, after estimating previous reflection coefficients k_1,… k_{p-1} through a recursive process, k_p reflection coefficient is determined.

At each level, reflection coefficient is calculated as bellow:

$$k_p = \frac{-2\sum_{t=p+1}^{N} e_{f,p-1}(t)\, e^*_{b,p-1}(t-1)}{\sum_{t=p+1}^{N}\left[\left|e_{f,p-1}(t)\right|^2 + \left|e^*_{b,p-1}(t-1)\right|^2\right]} \tag{3}$$

where, $e_{f,p-1}$ and $e_{b,p-1}$ are forward and backward forecasted errors for $(p-1)^{th}$ order of the model. In the present work, AR coefficients from different orders of Burg's method based on Levinson–Durbin recursion algorithm were extracted as feature vectors, and the results of classification accuracies were compared.

4. Raw Signal Features

Without any processing of signals, baseline features are extracted along with Hjorth parameters and Katz Fractal Dimension.

1. Hjorth Parameters

These are simple statistical features computed using the following expressions:

$$\text{Mobility: } \sqrt{\frac{variance(\dot{S}(t))}{variance(S(t))}},\ \text{Complexity: } \sqrt{\frac{mean(\dot{S}(t))}{mean(S(t))}}$$

An additional feature, 'Activity', was omitted because it is just square of the standard deviation (i.e. the variance) and the standard deviation is already included among the statistical features above.

2. Fractal Dimension

This feature also seeks to capture information about the shape of the signal. The formal way of defining dimension is to consider the scaling relationship between units of measurement and the number of such units required to measure a shape, as shown in the following equation.

$$N \propto \epsilon^{-D}$$

In the above ϵ denotes the amount by which the unit of measurement is increased or reduced, N denotes the number of the newly scaled units of measurement required to measure the same shape and D is the fractal dimension.

IV. FEATURE SELECTION

Dimensionality reduction is one of the basic problems in pattern recognition and classification. The main purpose of feature transformation and feature selection [4] is to remove redundant or irrelevant information and improve classification performance.

In this paper we use Fisher score ranking algorithm for selecting the top features.

A. *Fisher Score*

Fisher score is a kind of commonly used supervised feature selection algorithm. The basic idea is to find a feature subset, which satisfies maximizing between-class distances and minimizing within-class distances of samples. The expression of Fisher score is shown below.

$$F(\boldsymbol{Z}) = tr\left\{\boldsymbol{S}_b \boldsymbol{S}_w^{-1}\right\}$$

where*tr{.}* denotes the trace of matrix. The between-class scatter matrix $S_h$and within-class scatter matrix S_w are defined as

$$\boldsymbol{S}_b = \sum_{i=1}^{c} n_i (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^{\mathrm{T}}$$

$$\boldsymbol{S}_w = \sum_{i=1}^{c}\sum_{j=1}^{n_i} (z_{ij} - \boldsymbol{\mu}_i)(z_{ij} - \boldsymbol{\mu}_i)^{\mathrm{T}}$$

whereμ is average value of all samples, μ_i is average value of samples in the i^{th} class, $n^i$is the number of samples in the i^{th} class, and c is the number of classes. In order to facilitate solving the algorithm, feature selection based on Fisher score can also be formulated as a global optimization problem:

$$\begin{aligned}&\max\ \operatorname{tr}\left\{\left(\operatorname{diag}(\boldsymbol{w})\boldsymbol{S}_b\operatorname{diag}(\boldsymbol{w})\right)\left(\operatorname{diag}(\boldsymbol{w})\boldsymbol{S}_w\operatorname{diag}(\boldsymbol{w})\right)^{-1}\right\}\\ &\text{s.t. } \boldsymbol{w}_i \in \{0,1\},\ i=1,\ldots,d,\ \|\boldsymbol{w}\|_1 = m\end{aligned}$$

wherediag($\boldsymbol{w}$) is a diagonal matrix and $\boldsymbol{w}$ is its primary diagonal element.

V. TRAINING AND CLASSIFICATION

Artificial Neural Network (ANN), a paradigm that is related to biological networks and tries to mimic the structure of the human brain. A neural network is a massively equivalent distributed process, made up of simple processing units, which has a property for storing knowledge and making it available for use. One of the most important properties of neural networks is their ability to learn from examples, that is, learn to produce a certain output when fed with a certain input. The learning process involves modification of the connection weights, to make its overall performance correspond to a desired performance defined by the set of training examples. For each example in the training set there exists an input pattern and a desired output pattern. To train the network an example from the training set is chosen and fed to the network to see what output it produces. If the expected output is not obtained, the internal weight of the network is modified according to some training algorithm, so as to minimize the difference between the desired and the actual output. The training is then continued with another training example and so on, until the network has reached steady state. Here a fully connected network is employed and the standard back propagation algorithm can be used for training.

VI. RESULTS AND EXPERIMENT

The experimentation begins with pre-processed SEED data set by Clipping and Scaling them to custom low threshold values. The reason for clipping the signal is due to the fact that the variation in low amplitude values of EEG signal consists of the most important information in emotion classification, which has been found with repeated experiments.

With this clipped dataset, Hybrid Feature Extraction procedure is followed. Starting with Raw signal, DWT decomposition, WPT decomposition until Auto Regression, the clipped datasets are transformed and the baseline features are calculated for the transformed signals. These features that are extracted from each and every transformed signal is fused together as single list of features.

Fisher scoring algorithm is applied on the fused feature list to determine the top features in it, which returns list of feature identifiers and their corresponding scores. By using the scores obtained, top 3000 features having high scores is selected.

The selected 3000 features are then passed into ANN for training using scaled conjugate gradient backpropagation algorithm.

After successful training, the trained Neutral Network is tested with a total of 675 samples from SEED dataset. The performance metrics of the testing are as below.

1. *Confusion Matrix*

Higher green % value is better.

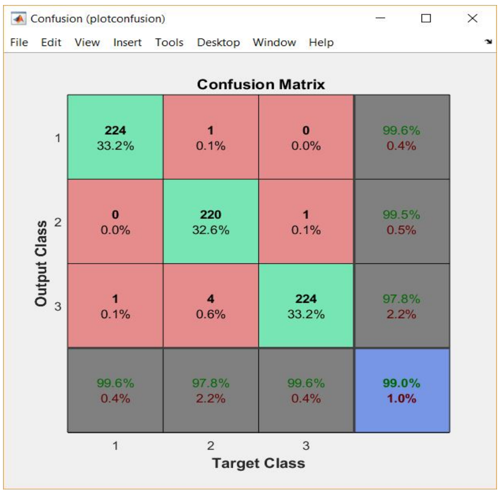

Fig. 3 Hybrid Extraction Confusion Matrix for testing

2. *Cross Entropy Performance*

Lower value is better.

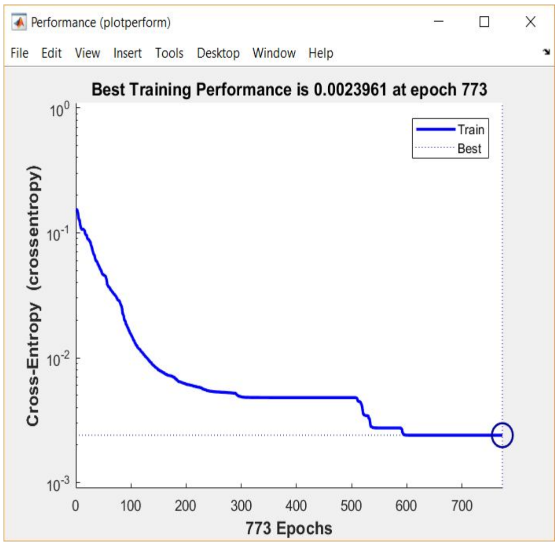

Fig. 4 Hybrid Extraction Cross entropy performance for testing

3. *Receiver Operating Characteristics (ROC)*

Closer to the top left corner is better.

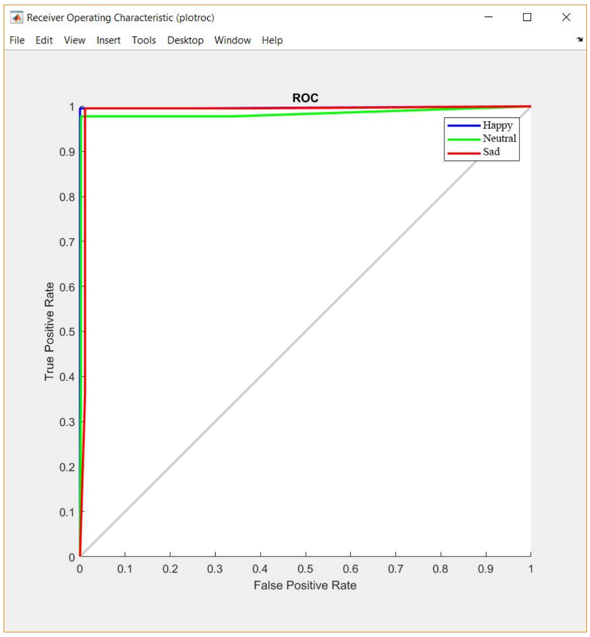

Fig. 5 Hybrid Extraction ROC for testing

VII. CONCLUSION

In this paper, the Hybrid Feature Extraction technique is mainly focused for accuracy, Fisher Scoring algorithm is used for Feature Selection process and Artificial

Neural Network is used for Classification. Initial Clipping and Scaling of data proves to be efficient for gaining higher accuracy. It is also considered that the feature extraction and training process did not require unacceptable amount of time.
For classification into 3 classes: Positive Emotion, Neutral Emotion and Negative Emotion, this approach classified 99% of 675 sample test data correctly.

REFERENCES

[1] Wei-Long Zheng and Bao-Liang Lu, “A multimodal approach to estimating vigilance using EEG and forehead EOG”., Journal of Neural Engineering, Vol. 14, Issue: 2, pp. 026-017, 2017.

[2] E. Gallego-Jutgla, . Sole-Casals, F. B. Vialatte, M. Elgendi, A. Cichocki, and J Dauwels, "A hybrid feature selection approach for the early diagnosis of Alzheimer's disease", Neural Engnieering, Vol. 12, Issue: I, pp. 016-018, 2015.

[3] F. Lotte, M. Congedo, A. Lecuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces ", I. Neural Engineering, Vol. 4, Issue: 2, pp. RI-RI3, 2007.

[4] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from EEG", IEEE Transactions on Affective Computing, Vol. 5, Issue: 3, pp. 327-339, 2014.

[5] Mohammad Soleymani, SadjadAsghari-Esfeden, Yun Fu, “Analysis of EEG Signals and Facial Expressions for Continuous Emotion Detection”, IEEE Transactions on Affective Computing Vol: 7, Issue: 1, Jan.-March 1 2016.

[6] U. R. Acharya, R. Yanti, G. Swapna, V S. Sree, R. J Martis, d J S. Suri, "Automated diagnosis of epileptic electroencephalogram using independent component analysis and discrete wavelet transform for different electroencephalogram durations", Proceedings of Institution of Mechanical Engineers Vol. 227, Issue. 3, pp. 234-244, 2013.

[7] H. Joho, J. Staiano, N. Sebe, and J. Jose, “Looking at the viewer: analyzing facial activity to detect personal highlights of multimedia contents”, Multimedia Tools and Applications, Vol. 51, Issue: 2, pp. 505–523, 2010.

[8] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from EEG", IEEE Transactions on Affective Computing,Vol. 5, Issue: 3, pp. 327-339, 2014.

[9] J.Atkinson, and D. Campos, "Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers", Expert System Application, Vol. 47, pp. 35-41, 2016.

[10] G. Rodriguez-Bermudez, P. J. Garcia-Laencina, .I. Roca-Gonzalez, and J. Roca-Dorda, "Efficient feature selection and linear discrimination of EEG signals ", Neuro computing, Vol. 115, pp. 161-165, 2013.

[11] Y. Song, and P. Lib, "A new approach for epileptic seizure detection: sample entropy based feature extraction and extreme learning machine", Journal of Biomedical Science and Engineering, Vol. 3, Issue: 6, pp. 556- 567, 2010.